Data Science Self-Study Resources

Galvanize Inc.

Contents

Have you been thinking about applying to Galvanize? How do you know you have enough of a technical background to be successful in our Data Science Immersive course? Find out by working through this document.

This document was designed to help you to gauge your level of understanding in programming (both Python and SQL) and to test your proficiency in Probability and Statistics. It also references resources which will help you along the way. If you want to learn more about the Admissions process itself, you will find a full description in DSI_Admission_process.pdf.

NOTE: This text can be shared publicly!

1 - Python

Python is an extremely popular and powerful programming language with tools to work in nearly every application area. In particular, Python (and specifically the Anaconda environment that supports it) is the industry standard for machine learning. As such, comfort with Python is a must

for any aspiring data scientist. In the Galvanize Data Science Immersive we teach data science using Python, and only briefly review python topics during the course. We therefore require that students pass a python coding interview.

How do you know you are ready?

You should be ready for the program if you are able to comfortably solve most CodeWars (http://bit.ly/2iWF0LL) Level 6-7 problems in about 30 minutes.

Resources

These are our top 3 online resources. They cover similar topics, but have their own approach to teaching Python. 1. Learn Python the Hard Way (Exercises 0-39): A combination of explanations and exercises to apply the concepts (https://learnpythonthehardway.org/book/) 2. Codecademy Python (Chapters 1-8, 12): An online interactive coding environment that provides lot of guidance and automatic feedback (https://www.codecademy.com/learn/python) 3. Google's Python Course (up to and including dictionaries): Nice resource if you like to have videos to follow along (https://developers.google.com/edu/python/)

Taking the Python workshop offered at Galvanize is a great option if you want to learn on campus. Even better, if you take the workshop and are accepted into the Galvanize Data Science program, we will credit the amount you paid for the workshop to your tuition balance. Galvanize Part-Time Course: Python Fundamentals (http://www.galvanize.com/courses/intro-to-python/)

Do you need more fluency?

Practice on coding challenge websites like Codewars (https://www.codewars.com/) and CoderByte (https://coderbyte.com). Use Stackoverflow when you're stuck. Keep practicing.

Want to practice on some mock challenges? Checkout the Python section of the self assessment document and work through the Python section of the coding challenges.

---

2 - SQL Resources

SQL is a building block of many things we do. If you are familiar with programming or data analysis, learning to write basic SQL queries won't take very long.

How do you know you are ready?

You should be familiar with SELECT, FROM, WHERE, CASE clauses and know how to use aggregates as well as JOINs!

Resources

There are many online resources to learning SQL. Here are some recommendations: 1. SQLZoo (Sections 0 to 7, http://sqlzoo.net/) 2. Codacademy SQL (Chapters 1-4, without the pro version, https://www.codecademy.com/learn/learn-sql)

Want to practice on some mock challenges? Checkout the SQL section of the coding challenges.

3- Statistics and Probability Resources

In the Galvanize Data Science Immersive, we review probability, statistics and classical regression, but then build on these topics as we learn advanced and modern machine learning methodology. It is therefore extremely important that our students be quantitatively and analytically literate, with a solid background in baseline probability and statistical theory. We evaluate technical analytical skills though a math interview focusing on probabilities, statistics and basic model building.

How do you know you are ready?

The math background we want to see will allow you to (1) comfortably solve probability problems, (2) leverage statistical distributions to solve real world problems, and (3) perform hypothesis testing. A solid understanding of basic tools for linear regression is also appreciated.

Resources

Never taken a stats course?

We recommend you take a full-length course in statistics or probability before applying. Khan Academy and Udacity provides some good "entry level statistics videos" and a good introductory course is available on Udacity. Make sure to work through the exercises and take the quizzes.

All of these resources cover the same basic information, just in different ways and formats. - Udacity: Intro to Statistics (https://www.udacity.com/course/intro-to-statistics--st101) - Khan Academy: Probability and Statistics (https://www.khanacademy.org/math/statistics-probability) - Coursera, edX, and other sites also have a large number of introductory statistics classes.

Want to refresh your knowledge?

These are areas you should focus on: - Counting: permutations, combinations - Probability: of a single event, of 2 or more events (conditional probability, independent and dependent events, mutually exclusive events, Bayes' Theorem) - Probability distributions: Binomial, Geometric, Poisson distributions; Uniform, Normal distributions - Hypothesis testing: one/two sided test, confidence intervals, inference for two means, inference for proportions, chi-square tests - Linear Regression: build simple models, model performance, bias/variance trade-off

There are excellent exercises mentioned in the resources above. Should you wish for some additional learning material online, we would recommend this Probability Review (http://www.intmath.com/counting-probability/counting-probability-intro.php) and this Statistics Videos (http://www.jbstatistics.com/).

Want to go further?

To go beyond the call of duty, and make sure you are in great shape for the program, work your way through chapters 1-4 of an Introduction to Statistical Learning. You could also enroll in this Stanford online course or find the same material in this In Depth Introduction to Machine Learning.

Want to practice on some mock challenges? Checkout the math challenges of the self assessment document.

---

4 - Comprehensive

Our Data Science Primer: https://github.com/zipfian/data-science-primer

This is the most comprehensive source and excellent preparation for the program. It is a substantial packet that may take some time. However, it is a great resource for potential students.